MINIMUM SUMS OF SQUARES WHEN THE TWO EXTREMES ARE RETAINED

BU-28-M Prasert Na Nagara June, 1952

In using Duncan's Multiple Comparisons Test, one has to find whether the sum of squares of each combination of the ranked means, chosen from $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ and including $\bar{x}_1$ and $\bar{x}_n$, exceeds a calculated significant magnitude. The work is much reduced, if one can find the combination which gives the smallest sum of squares among m varieties, because if this exceeds the calculated number then the rest of the combinations will also.

<u>Rules</u> 1. Coding the ranked means $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ (in ascending magnitude) by subtracting the mean of all means, $\bar{x} = \sum_{i=1}^{n} \bar{x}_i/n$, from each term, one gets the deviations from the overall mean, $y_1, y_2, \ldots, y_n$ whose sum is equal to zero.

2. To get the smallest sum of squares among n-1 means, one excludes either y_2 or y_{n-1} whichever has the greater absolute value.

3. To get the smallest sum of squares among n-2 means, one excludes either (a) y_2 and y_3, (b), y_2 and y_{n-1}, or (c) y_{n-2} and y_{n-1}. Between (a) and (b), one excludes the values in (a) if $2y_2 > (\sqrt{n-2}+1)(-y_{n-1}-y_3)$ and excludes the values in (b) otherwise. Between (c) and (b) one excludes the values in (c) if $-2y_{n-1} > (\sqrt{n-2}+1)(y_2+y_{n-2})$ and excludes the values in (b) otherwise.

4. Similarly, to get the smallest sum of squares among n-k means, one excludes a combination of k means on one or both extremes.

<u>Note</u> It is advisable to find $S = y_1^2 + y_2^2 + \ldots + y_n^2$ first. Then the sum of squares among n-1 means excluding y_i is easily shown to be $S - y_i^2 - y_i^2/(n-1)$ --(1), and the sum of squares among n-2 ranked means excluding y_i and y_j is $S - y_i^2 - y_j^2 - (y_i + y_j)^2/(n-2)$ -- (2) and so on.

Rule 2 is obvious from (1). A proof for rule 3 is sketched below.

From (2), it is known that the sum of squares of n-2 means will be a minimum when $M = y_i^2 + y_j^2 + (y_i + y_j)^2/(n-2)$ is a maximum. Among all pairs of positive y's (y_1 not included) y_2 and y_3 give a maximum M. Among all pairs of negative y's (y_n not included) y_{n-2} and y_{n-1} give a maximum M. If $y_3 > |y_{n-1}|$ then y_2 and y_3 certainly give a maximum M among all pairs. If $|y_{n-2}| > y_2$ then y_{n-1} and y_{n-2} give a maximum M.

Now, one needs to compare y_2, y_3 with y_2 and $y_{k>3}$. One excludes y_2 and y_3 if $y_2^2 + y_3^2 + (y_2 + y_3)^2/(n-2) > y_2^2 + y_k^2 + (y_2 + y_k)^2/(n-2)$

$$
\begin{aligned}
\text{or} \qquad (n-2)(y_3^2 - y_k^2) &> (y_2 + y_k)^2 - (y_2 + y_3)^2 \\
&= (y_k - y_3)(2y_2 + y_3 + y_k) \\
(n-2)(y_3 + y_k) &> -(2y_2 + y_3 + y_k) \text{ since } y_3 - y_k \text{ is}
\end{aligned}
$$

positive.

$$\text{Also,} \qquad 2y_2 > (n-1)(-y_k - y_3).$$

Since the right hand side is maximum when $k = n-1$, one needs to test y_2, y_3 with y_2 and y_{n-1} and rule 3 is proved.

Similarly, rule 4 may be proved and in fact among combinations of m means, one would exclude the one containing y_i instead of y_{i+j} if $2Y > (m+1)(-y_i - y_{i+j})$ and exclude y_{i+j} instead of y_i, otherwise. Here Y = the sum of all m-1 values of the y's excluded.

<u>Example</u> From the 9 ranked means 205, 200, 200, 185, 143, 140, 138, 133, and 132 one wants to find the combination of 8 means (including the first and the last means) which gives the minimum sum of squares. Coding by subtracting 164, the mean of 9 means, one obtains 41, 36, 36, 21, -21, -24, -26, -31, and -32 as y_1, y_2, ..., y_9. $\sum_1^9 y^2 = 8392 = S$. Since $y_2 > |y_8|$, the sums of squares among the 8 means excluding y_2, $ss_{13456789}$, gives the minimum sum of squares among any 8 means and $S - y_2^2 - y_2^2/(n-1) = 6934$.

Among the sums of squares of 7 means one should exclude either y_2, y_3; y_2, y_8 or y_7, y_8. Since y_2 and y_3 are both greater than y_7, y_8, it is obvious that one should exclude y_2, y_3, and the minimum sum of squares among 7 means is $ss_{1456789} = S - y_2^2 - y_3^2 - (y_2 + y_3)^2/(n-2) = 5059$.

To get the minimum sum of squares among 6 means one should exclude either (1) y_2, y_3, y_4; (2) y_2, y_3, y_8; (3) y_2, y_7, y_8; or (4) y_6, y_7, y_8. Now, $2(y_2 + y_3) > (6 + 1)(-y_4 - y_8)$. Therefore, if the values in (1) are excluded one gets a smaller sum of squares than if the values in (2) are excluded. Since $2(y_2 + y_8) > (6 + 1)(-y_3 - y_7)$ the values in (2) are excluded instead of the values in (3). Similarly, (4) is preferable over (3). One needs only to compare (1) with (4).

$$y_2^2 + y_3^2 + y_4^2 + (y_2 + y_3 + y_4)^2/(n-3) = 4475 > y_6^2 + y_7^2 + y_8^2 + (y_6 + y_7 + y_8)^2/(n-3) = 3307.$$

Therefore, excluding y_2, y_3, y_4 one gets the minimum sum of squares, $ss_{156789} = 8392 - 4475 = 3917$.